DOI:10.19651/j.cnki.emt.2416511

基于全景分割与多视图几何的动态 SLAM 方法*

王 爽[1] 刘云平[1] 张柄棋[1] 陆旭春[1] 徐 梁[2]
(1.南京信息工程大学自动化学院 南京 210044; 2.中国人民解放军海军大连舰艇学院 大连 116018)

摘 要: 在 SLAM 系统估计相机位姿时,大量运动物体的特征点参与特征跟踪线程导致算法准确性和鲁棒性下降,因此如何高效准确地剔除场景中的动态物体尤为重要。现有的动态视觉 SLAM 算法在处理动态物体时可能漏检或是错误地将静态物体识别为动态物体并将其剔除,引发静态特征点数量不足的问题,进而影响 SLAM 系统的稳定性和精度。因此提出一种基于全景分割与多视图几何的视觉 SLAM 方法,该算法使用全景分割 FPN 网络准确识别分割图像中的所有物体,剔除先验动态特征点并尽可能多地保留静态特征,在此基础上使用融合图像金字塔的 LK 光流法实现光流跟踪并剔除平行动态特征点,潜在的动态特征点则采用基于动态概率的多视图几何法更有效地对其剔除,避免了动态特征点漏检的情况,实现对场景中动态物的全面筛查以提高系统精度。在系统构建的稀疏点云的基础上实现对语义地图与八叉树地图的构建。实验使用 TUM RGB-D 数据集验证系统定位精度,结果表明,与 ORB-SLAM2 相比,本算法在所有序列的绝对轨迹误差的均方根误差(RMSE)平均降低了 84.34%,显著提升了系统的鲁棒性和准确性,并且构建两种可用于 SLAM 上层任务的地图,具有一定的使用价值。

关键词: 视觉 SLAM;动态场景;全景分割;多视图几何;图像金字塔;八叉树地图
中图分类号: TP391.9;TN98 **文献标识码**: A **国家标准学科分类代码**: 510.99

Dynamic SLAM approach based on panoptic segmentation and multi-view geometry

Wang Shuang[1] Liu Yunping[1] Zhang Bingqi[1] Lu Xuchun[1] Xu Liang[2]
(1. School of Automation, Nanjing University of Information Science and Technology, Nanjing 210044, China;
2. PLA Dalian Naval Academy, Dalain 116018, China)

Abstract: When the SLAM system estimates the camera position, a large number of feature points of moving objects participate in the feature tracking thread leading to a decrease in the accuracy and robustness of the algorithm, so how to efficiently and accurately reject the dynamic objects in the scene is particularly important. Existing dynamic vision SLAM algorithms may miss detecting or incorrectly recognize static objects as dynamic objects and reject them when dealing with dynamic objects, which triggers the problem of insufficient number of static feature points, thus affecting the stability and accuracy of the SLAM system. Therefore, this paper proposes a visual SLAM method based on panoptic segmentation and multi-view geometry, which uses panoptic segmentation FPN network to accurately recognize all objects in the segmented image, rejects a priori dynamic feature points and retains as many static features as possible, based on which LK optical flow method with fused image pyramid is used to realize optical flow tracking and reject parallel dynamic feature points, and potential dynamic feature points are used to track the dynamic feature points. The potential dynamic feature points are rejected more effectively by the multi-view geometry method based on dynamic probability, which avoids the omission of dynamic feature points and realizes the comprehensive screening of dynamic objects in the scene to improve the accuracy of the system. The construction of semantic map and octree map is realized on the basis of sparse point cloud constructed by the system. The experiments use the TUM RGB-D dataset to verify the system localization accuracy, and the results show that the root mean square error (RMSE) of the absolute trajectory error of this algorithm is reduced by an average of 84.34% in all sequences compared with ORB-SLAM2, which significantly improves the robustness and accuracy of the system, and it is of use to construct two maps that can be used for SLAM upper layer tasks.

Keywords: visual SLAM;dynamic scenes;panoptic segmentation;multi-view geometry;image pyramids;octree maps

0 引 言

视觉同时定位与建图[1](visual simultaneous localization and mapping,VSLAM)在自动驾驶、无人机导航、仓储物流及智能家居等领域展现出广泛的应用潜力,但其性能在动态环境中常常受限。这是因为大多数现有 VSLAM 算法基

收稿日期:2024-07-24
* 基金项目:空间智能控制技术重点实验室稳定支持基金(HTKJ2023KL502020)、江苏省现代农机装备与技术示范推广项目(NJ2023-19)、江苏省农业科技自主创新资金(CX(23)3143)项目资助

于静态环境的假设[2]，而现实世界中动态物体的存在对这一假设构成了挑战，尤其在城市巷战、无人驾驶、仓储物流等场景下，动态物体占据了多数，若将它们假设为静态背景或静态物体，则不可避免地对 SLAM 系统造成影响，并且系统的误差也会随着动态物的增多而增加。

为提升动态环境下 VSLAM 准确性，需要机器人能够识别和区分出环境中的动态或潜在动态对象。随着深度学习的飞速发展以及计算机算力的不断提升，这一任务已变得更加可行。因此，出现了许多结合深度学习技术来剔除动态目标，从而基于静态背景进行精确 SLAM 的算法[3]。国内外的研究学者在此领域上进行深入研究并取得了一系列的成果。清华大学的研究者提出了 DS-SLAM 算法[4]，该算法利用 SegNet 进行语义分割，并结合极线检测结果来识别并剔除动态目标，显著提高了动态场景下的 SLAM 精度，但其会误剔除静态特征点，从而导致特征点跟踪失败。此外，Bescós 等[5]提出了 DynaSLAM 算法，该算法不仅依赖深度学习技术识别潜在动态目标，还结合了多视图几何法来检测没有先验信息的动态区域，但该算法并没有对场景中运动方向与相机运动方向平行的动态特征进行检测，降低了系统定位的精度。Cui 等[6]提出了 SOF-SLAM 算法，该算法使用 SegNet 网络实现语义分割，根据光流法特征匹配的结果计算图像帧间的基础矩阵，再结合几何运动判别动态点，但该算法采用原始光流法容易受环境噪声影响，同时 SegNet 难以完整分割图像中的所有物体及背景。中国科学院大学的马萍[7]则结合了 Mask R-CNN 实例分割网络，对关键像素点进行运动一致性检测以判断其真实状态，从而去除动态点，但该算法通过计算两帧间同一个特征点的角度变化判断其是否属于动态特征，而跟踪过程中特征点角度的变化对低动态物体不敏感，并且实例分割在面对大视野范围里存在的动态特征点分割效果较差。韩国科学技术院的 Song 等[8]针对动态环境下的错误回环问题，提出了 DynaVINS 算法，该算法结合了 IMU 观测并设定关键帧组来减少静态目标对建图的影响。Liu 等[9]提出了 RDS-SLAM 算法，该算法提出了一种新的基于语义的实时动态 vSLAM 算法，并采用一种关键帧选择策略，尽可能使用最新的语义信息。多伦多大学太空研究所的 Qian 等[10]则通过概率估计跟踪目标状态，并根据稳定性评分来建模和维护更稳定的静态地图。上述 3 种算法仅构建了稀疏点云地图，难以满足更高级的导航、规划等任务。Liu 等[11]提出的 RDMO-SLAM 通过添加基于稠密光流的语义标签预测利用更多的语义信息，并将地图点的运动速度作为约束条件来减少动态特征点对系统的影响。

综上所述，针对现有的动态视觉 SLAM 算法在动态环境下难以完整识别动态物、难以准确判断出平行动态特征与潜在动态特征[12]，从而出现漏检或误检场景中动态物体并无法构建准确的场景语义与八叉树地图的问题[13]，提出了一种基于全景分割与多视图几何的视觉 SLAM 算法。本文的主要创新点如下：

1）为提高识别场景中动态物体的精度和效率，采用 Panoptic FPN（panoptic feature pyramid networks）全景分割对图像进行预处理，该方法对场景中所有物体及背景进行了像素级的分割，在准确高效地标记并剔除动态特征点的同时尽可能多地保留静态物体及静态背景。

2）为避免任务场景中出现与相机运动方向平行的动态特征点对 SLAM 系统造成影响，提出了一种基于图像金字塔的 LK（lucas kanade）光流法，并利于得到的光流值设计了一种平行动态特征点的剔除策略。

3）为进一步提高系统在动态场景中的定位精度，提出了一种融合了前一帧动态概率的多视图几何方法对场景中潜在的动态特征点进行剔除，进一步保证了系统精度。

4）针对稀疏点云地图无法执行 SLAM 上层任务的弊端，设计了局部建图线程，提供了语义地图与八叉树地图两个选项。语义地图提供了人机交互的可能，场景的三维重建则依赖于八叉树地图。

1　算法整体流程框架

本研究提出的算法在 ORB_SLAM2[14]的基础上进行改进，作为一套经典的基于特征点的视觉 SLAM 算法，ORB-SLAM2 包含跟踪、局部建图和闭环 3 个核心线程[15]，但并没有加入处理动态物体的方法。因此本文在此基础上添加了一系列剔除动态特征点的策略，改进后算法的工作流程如图 1 所示。在跟踪线程前引入了图像预处理线程，该线程采用先进的全景分割技术[16]，预先获取图像中的语义信息并筛选先验动态特征，将其剔除，从而为提高算法的稳定性和准确性打下基础。此外，在跟踪线程中增加了平行动态特征点和潜在动态特征点的筛选剔除策略。在局部建图线程中，结合了从全景分割网络中获得的丰富语义信息和 SLAM 系统获得的稀疏点云数据[17]，以实现更为精细和准确的语义度量地图和八叉树地图构建[18]。

在数据预处理线程中使用 Panoptic FPN 全景分割对图像进行预处理，将像素分类为 Things 或 Stuff。Things 是具有明确边界的可数对象，并且可能是可移动的，例如人、动物或车辆。Stuff 是指图像中不可数的无定形区域，大多是“不可移动的”，如天空、地板或墙壁。该线程将标注为 Things 的物体进行剔除，保留 Stuff 物体，将分割后的图像输入到跟踪线程中。

跟踪线程中使用融合了图像金字塔的 LK 光流法对保留下的像素点进行跟踪，得到图像中每个像素的光流值，使用设计的结合 Z-score 和距离的方法判断场景中运动方向与相机运动方向平行的动态物体，检测将其剔除。

经过上述方法，已对图像中大部分的动态物体进行检测剔除处理，但场景中仍会出现因人为因素而产生运动的物体需要进一步剔除，提取完 ORB 特征点后，因此采用一种融合前一帧动态概率的多视图几何方法对潜在动态特征

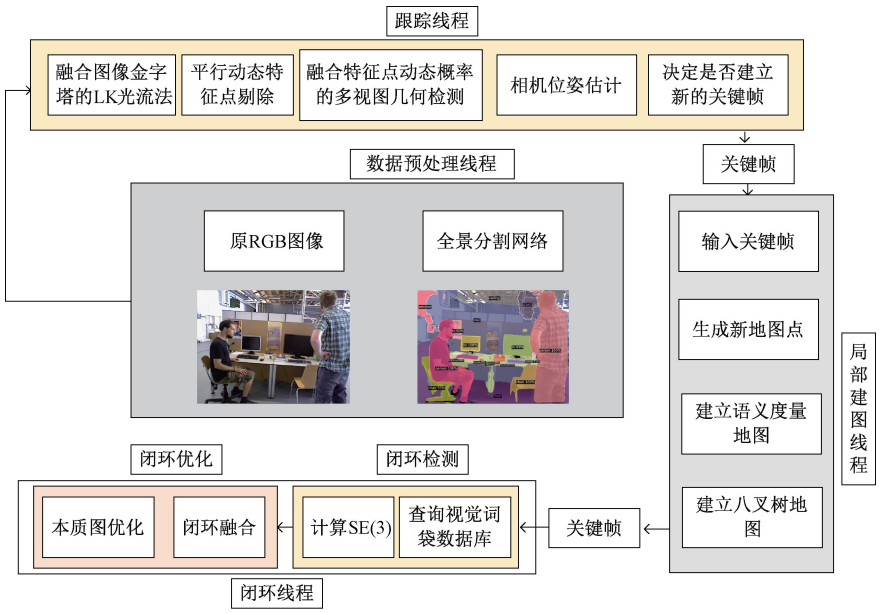


图 1　本文算法流程图

Fig. 1　Algorithm flowchart of this paper

点进行剔除,仅使用剩下的静态特征点进行相机位姿估计。局部建图线程中使用筛选出的关键帧,生成新的地图点,建立语义度量地图与八叉树地图。最后,闭环线程通过词袋查询关键帧数据库,找到和当前关键帧可能发生闭环的候选关键帧,求解它们之间的 Sim(3)变换,最后执行闭环融合和本质图优化,优化所有关键帧的位姿。

2　基于全景分割的动态点检测

2.1　Panoptic FPN 全景分割

未经处理的图像往往存在一些较为明显的动态物体,为了快速准确地识别这些影响定位精度的动态物体,本文在数据预处理线程中使用 Panoptic FPN 全景分割网络进行图像分割,该网络在 Mask R-CNN 的基础上进行了扩展,增加了一个使用 FPN 为主干的语义分割分支,实现了全景分割的任务。全景分割网络不仅包含了实例分割的功能,能够识别并定位图像中的个体对象,还具备了语义分割的能力,可以对图像中的每个像素进行分类。Panoptic FPN 框架示意图如图 2 所示。

Panoptic FPN 的核心在于其主干网络——特征金字塔网络(FPN),它利用了标准的多尺度特征提取网络,并通过一个自上而下的通路结构进行特征融合。这个通路从网络的最深层开始,通过逐步上采样,并与较低层的特征进行相加,从而构建了一个具有原始图像四种不同分辨率(1/32、1/16、1/8、1/4)的特征金字塔。每个金字塔层级的通道维度保持一致,确保信息的连贯性和丰富性。

Panoptic FPN 的实例分割部分基于 Mask R-CNN 框架。在获得 FPN 提取的特征后,它利用基于区域的对象检测器,在特征金字塔的不同层级上执行 RoI 池化操作。

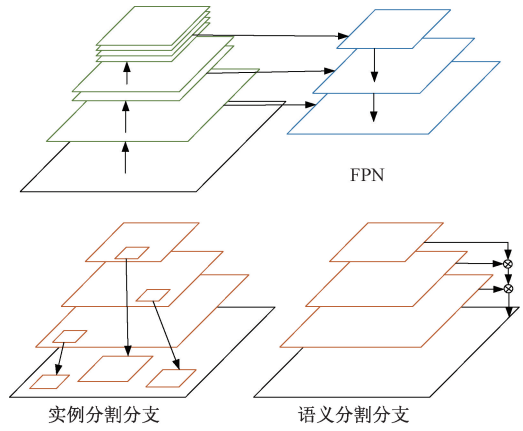


图 2　Panoptic FPN 框架

Fig. 2　Framework of Panoptic FPN

这一过程允许网络在多个尺度上识别并定位图像中的个体对象。此外,Mask R-CNN 还增加了一个网络分支,用于预测每个感兴趣区域(ROI)的类别标签、边界框以及像素级的二进制分割掩码。

至于语义分割部分,Panoptic FPN 采用了轻量级的密集预测分支。这个分支在特征金字塔的 4 个尺度(1/32、1/16、1/8、1/4)上分别执行不同次数的上采样操作(分别为 3 次、2 次、1 次、0 次),以获得与原始图像 1/4 分辨率相匹配的特征图。随后,这些特征图通过按元素相加的方式进行融合,最后经过卷积层、双线性上采样和激活函数,输出每个像素的预测类标签,实现图像的语义分割。

全景分割网络的分割结果如图 3 所示。图中将所有的像素进行了分割,并将它们分为 Stuff 和 Things 两类,在数据预处理阶段仅剔除 Stuff(移动的人),而将 Things(场

景背景)保留为后续流程提供静态特征点。

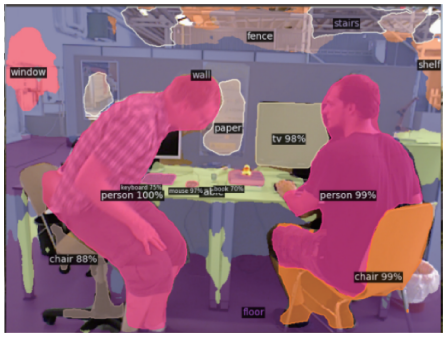


图 3　全景分割结果

Fig. 3　Result of panoptic segmentation

2.2　融合图像金字塔的 LK 光流法

LK 光流法通过计算像素在两帧间的运动变化从而实现对图像中物体的跟踪,但由于像素的非线性非常强,在优化光流值的过程中,若求解出的光流值离真正的最优解相差 10 个像素时,那么在优化函数迭代求解的过程中很容易陷入局部最小。因此本文提出通过使用图像金字塔来提升光流跟踪的稳定性。图像金字塔示意图如图 4 所示。

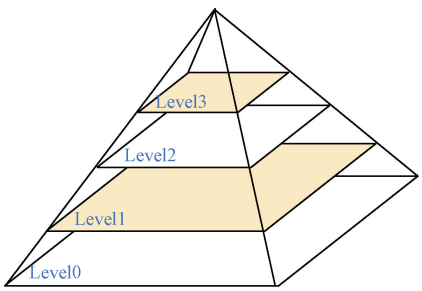


图 4　图像金字塔示意图

Fig. 4　Image of pyramid schematic

图像金字塔中 level0 和 level3 之间的缩放比例为 1:2,那么在 level0 上光流求解值和最优值相差 10 个像素的距离经过缩放后像素距离可降为 5 个像素,因此在缩放过后的图像上更容易成功跟踪。当然如果只用缩放后的图像进行跟踪是不可行的,仍然需要回到原来的原始图像上,在缩放后的图像上若是 2 个像素的置信度,则回到原图是 4 个像素的置信度,这会带来精度下降的问题。因此,对于每张待处理的图像,将其先缩放到 level3,在 level3 进行光流追踪,将追踪到的像素点恢复到 level2,作为 level2 追踪的初始值,如此反复,一直追踪到 level0 层。如此做不仅很大程度地避免了优化求解过程中陷入局部最小值,又避免了像素精度下降的问题,从而得到了更为精确的光流值。

光流法的成立基于以下 3 个假设[19]:

假设 1)灰度不变假设:同一个地图点的像素灰度在不同图像帧里的灰度是不变的。

假设 2)时间持续性假设:相机的的位姿随时间的变化是渐进而非突变的,相邻时间点的图像之间,像素灰度值不会发生剧烈变化。

假设 3)空间一致性:场景中属于同一表面的相邻点,在三维空间中具有相似的运动特性,这种相似性在它们被投影到图像平面上时表现为这些点在图像上的位置相互接近。

基于 A 和 B 假设,得到图像的约束方程:

$$I(x+\mathrm{d}x,y+\mathrm{d}y,t+\mathrm{d}t)=I(x,y,t) \tag{1}$$

$I(x,y,t)$ 表示在 t 时刻像素坐标为 (x,y) 位置的像素灰度值,$I(x+\mathrm{d}x,y+\mathrm{d}y,t+\mathrm{d}t)$ 表示在 $t+\mathrm{d}t$ 时刻像素坐标为 $(x+\mathrm{d}x,y+\mathrm{d}y)$ 位置的像素灰度。对式(1)进行泰勒展开,保留一阶项,得:

$$I(x+\mathrm{d}x,y+\mathrm{d}y,t+\mathrm{d}t)\approx I(x,y,t)+\frac{\partial I}{\partial x}\mathrm{d}x+\frac{\partial I}{\partial y}\mathrm{d}y+\frac{\partial I}{\partial t}\mathrm{d}t \tag{2}$$

由于灰度不变假设,因此 $t+\mathrm{d}t$ 时刻与 t 时刻的灰度值相同,从而得到:

$$\frac{\partial I}{\partial x}\mathrm{d}x+\frac{\partial I}{\partial y}\mathrm{d}y+\frac{\partial I}{\partial t}\mathrm{d}t=0 \tag{3}$$

两边同时除以 $\mathrm{d}t$,得:

$$\frac{\partial I}{\partial x}\frac{\mathrm{d}x}{\mathrm{d}t}+\frac{\partial I}{\partial y}\frac{\mathrm{d}y}{\mathrm{d}t}=-\frac{\partial I}{\partial t} \tag{4}$$

其中,$\mathrm{d}x/\mathrm{d}t$ 为像素在 x 方向上的速度,$\mathrm{d}y/\mathrm{d}t$ 为像素在 y 方向上的速度。$\partial I/\partial x$ 为图像在该点 x 方向上的梯度,$\partial I/\partial y$ 为图像在该点 y 方向上的梯度,记 $\mathrm{d}x/\mathrm{d}t$ 为 u,$\mathrm{d}y/\mathrm{d}t$ 为 v,$\partial I/\partial x$ 为 I_x,$\partial I/\partial y$ 为 I_y,把图像灰度对时间的变化量记为 I_t,则式(4)可写成:

$$\begin{bmatrix}I_x & I_y\end{bmatrix}\begin{bmatrix}u\\v\end{bmatrix}=-I_t \tag{5}$$

基于假设 B,在图像上设定一个 4×4 的窗口,4×4 窗口中的所有像素有着同样的运动,即 u,v 均相同。因此式(5)方程的求解可拓展为:

$$\begin{bmatrix}I_x & I_y\end{bmatrix}_{16}\begin{bmatrix}u\\v\end{bmatrix}=-I_{t16} \tag{6}$$

则记:

$$\boldsymbol{A}=\begin{bmatrix}I_x & I_y\end{bmatrix}_{16}=\begin{bmatrix}\begin{bmatrix}I_x & I_y\end{bmatrix}_1\\ \vdots \\ \begin{bmatrix}I_x & I_y\end{bmatrix}_{16}\end{bmatrix},\boldsymbol{b}=\begin{bmatrix}I_{t1}\\ \vdots \\ I_{t16}\end{bmatrix} \tag{7}$$

即:

$$\boldsymbol{A}\begin{bmatrix}u\\v\end{bmatrix}=-\boldsymbol{b} \tag{8}$$

式(8)是超定方程,则可求出此方程的最小二乘解为:

$$\boldsymbol{M}_i=\begin{bmatrix}u\\v\end{bmatrix}^*=-(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})^{-1}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{b} \tag{9}$$

得到 $\boldsymbol{M}_i$,其中,$\boldsymbol{M}_i$ 为第 i 个像素的光流值。

2.3　平行动态特征点剔除方法

场景中存在的运动方向与相机相同的平行动态特征点仍会对系统造成干扰,因此提出一种剔除平行动态特征点的方法。单张图片计算完所有静态像素的光流值后,得到了每一个像素点的光流值 $\boldsymbol{M}_i$。然后采用结合 Z-score 和距离的方法判断平行动态特征点。公式如下:

$$\boldsymbol{Z}_i = \frac{\boldsymbol{M}_i - \boldsymbol{M}_{ave}}{\boldsymbol{M}_{std}} \tag{10}$$

其中,$\boldsymbol{Z}_i$ 表示每个光流值的 Z-score 得分,$\boldsymbol{M}_i$ 表示单张图片第 i 个像素的光流值,$\boldsymbol{M}_{ave}$ 和 $\boldsymbol{M}_{std}$ 分别表示单张图片所有光流的平均值和标准差。

接下来,计算每个光流值与其他所有光流值的平均距离 $AvgDist_i$,公式如下:

$$AvgDist_i = \frac{1}{n-1}\sum_{j\neq i}\sqrt{(\boldsymbol{M}_i - \boldsymbol{M}_j)^2} \tag{11}$$

其中,n 是单张图片光流值的数目,$\boldsymbol{M}_j$ 表示单张图片第 j 个像素的光流值。

为了更准确地筛选出平行动态特征点,将 Z-score 和平均距离结合起来,得到:

$$CombinedScore_i = |\ Z_i\ | \times AvgDist_i \tag{12}$$

$$IsDyna_i = \begin{cases} True, & CombinedScore_i > T \\ False, & \text{其他} \end{cases} \tag{13}$$

其中,$CombinedScore_i$ 表示第 i 个像素的联合得分,设定阈值 T,若第 i 个像素的联合得分超过 T,则当前像素的判断值 $IsDyna_i$ 为 $True$,将当前像素视为动态特征点,剔除;反之,若 $IsDyna_i$ 为 $False$,则当前像素为静态特征点,保留。

2.4　基于动态概率的多视图几何方法

被人碰过的书和被人坐过的椅子仍可能移动,将其视为潜在动态特征点。因此提出了一种融合动态概率的多视图几何的判别方法进一步进行动态特征的检测剔除。

经过前两步的剔除策略,大部分动态特征点已被剔除,使用保留下来的静态部分进行 ORB 特征点的提取匹配。将匹配结果用于基础矩阵 $\boldsymbol{F}_{21}$ 的计算,这样可以提前避免一些不必要的动态点对计算基础矩阵造成影响,提高基础矩阵的可靠性,从而使得在进行几何法剔除动态特征点时获得更为可靠的先验信息,提高动态特征点剔除的准确性。

选取两帧间 8 对匹配点,利用归一化后的特征点坐标求取基础矩阵 $\boldsymbol{F}_{21}$,$\boldsymbol{F}_{21}$ 为当前帧到参考帧的基础矩阵。两帧之间的对极约束示意图如图 5 所示。

图 5 中,两个平行四边形表示着相机在两个位置的成像平面,分别为图像平面 1 和图像平面 2。O_1、O_2 分别表示相机在两个位置的的光心位置,它们两者之间的连线称为基线。基线与两个成像平面之间的交点 e_1 和 e_2 叫做极点,点 p 是空间中的三维地图点,该点在两个成像平面上的像素投影点分别为 $\boldsymbol{p}_1$ 和 $\boldsymbol{p}_2$,设 $\boldsymbol{p}_1 = [u_1, v_1, 1]$,$\boldsymbol{p}_2 = [u_2, v_2, 1]$,点 p、点 O_1 和点 O_2 三者一起形成的平面定义为为极平面。极平面分别和两个成像平面之间相交的线 $\boldsymbol{l}_1$ 和 $\boldsymbol{l}_2$ 称之为极线。

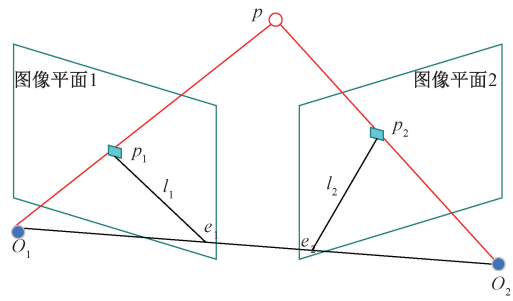


图 5　对极几何示意图

Fig. 5　Schematic diagram of the antipodal geometry

采用基础矩阵 $\boldsymbol{F}_{21}$ 进行几何约束:

$$\boldsymbol{p}_2^{\mathrm{T}}\boldsymbol{F}_{21}\boldsymbol{p}_1 = 0 \tag{14}$$

$$\boldsymbol{l}_2 = \boldsymbol{F}_{21}\boldsymbol{p}_1 = [A_2, B_2, C_2]^{\mathrm{T}} \tag{15}$$

$$\boldsymbol{l}_1 = \boldsymbol{p}_2^{\mathrm{T}}\boldsymbol{F}_{21} = [A_1, B_1, C_1] \tag{16}$$

此时,计算 $\boldsymbol{p}_2$ 点到极线 $\boldsymbol{l}_2$ 的距离:

$$d_2 = \frac{\boldsymbol{p}_2^{\mathrm{T}}\boldsymbol{l}_2}{\sqrt{\|A_2\|^2 + \|B_2\|^2}} = \frac{A_2u_2 + B_2v_2 + C_2}{\sqrt{\|A_2\|^2 + \|B_2\|^2}} \tag{17}$$

为了进一步检测动态特征点,提出了一种融合动态概率的多视图几何的判别方法,本发明结合前一帧的特征点动态概率来判断是否为动态特征点。公式如下:

$$P_{ci} = \frac{d_{2i}}{\max d_2} \times 100\% + \omega P_{li} \tag{18}$$

其中,d_{2i} 表示当前帧中第 i 个特征点到极线的距离,$\max d_2$ 为该距离中的最大值;P_{ci} 为当前帧中第 i 个特征点的动态概率,P_{li} 为前一帧中对应匹配点的动态概率;ω 是系数因子,本文取 0.5。当计算得到动态概率大于设定阈值时,本发明阈值设定为 80 %,将该特征点视为动态点;否则,当计算出的的距离小于设定阈值时,认为该特征点为静态特征点。

3　实验结果与分析

本文采用了 TUM RGB-D[20] 数据集来验证 RGB-D 场景下本文算法的定位精度实验效果。通过与 ORB-SLAM2 及其他相关的动态视觉 SLAM 进行对比,验证本文算法在动态场景中定位的准确性与鲁棒性。同时验证在构建场景语义度量地图与八叉树地图时,不会出现动态物体的点云数据与语义信息。本文算法运行环境为 Ubuntu20.04 操作系统,处理器为 Intel i7-12700H,显示适配器为 NVIDIA GeForce RTX 3050Ti GPU、Cuda 版本为 11.4。

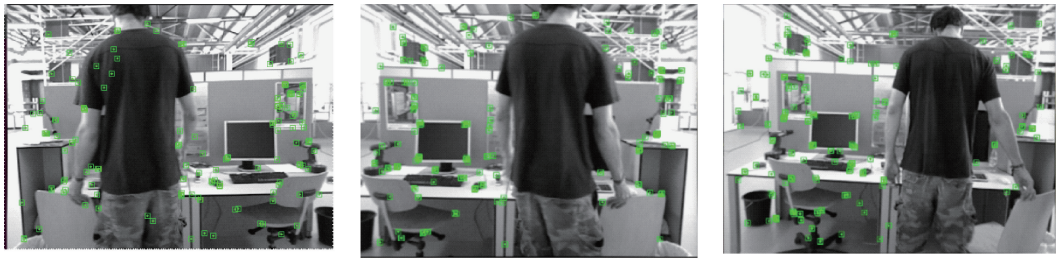
3.1　TUM 数据集

在本文的研究中,采用了 TUM RGB-D 数据集,该数

据集由 5 个室内环境下的 RGB-D 视频序列组成，每个序列均配备了高精度运动捕捉系统提供的真实轨迹数据。尤为重要的是，数据集中特别标记有 walking 序列用于评估视觉 SLAM 算法在复杂室内环境中处理中速移动动态对象的能力。这些 walking 序列具体展示了两个人在桌子周围持续走动的场景，其中动态对象长时间且显著地占据了相机视野的核心区域，为 SLAM 系统提出了严峻的挑战，要求算法在动态干扰下仍能维持稳定的定位和建图性能。

3.2　动态特征点剔除效果

图 6 中图像来源于 walking_xyz 序列，该序列中的人活动频繁，此时桌前的人将椅子挪开，人和椅子均处于运动状态。图 6(a)展示了 ORB-SLAM2 算法对数据集中特征点提取的结果，可以看出无论是动态还是静态物体均被提取。图 6(b)则展示了本文算法仅采用全景分割网络剔除动态特征点，虽然没有在人所在区域提取特征点，但是被人触碰过的椅子仍属于潜在动态物体而其动态特征点并未被剔除。图 6(c)则是本文算法同时使用全景分割、平行动态特征点剔除策略以及潜在动态特征点剔除策略共同作用的结果，可以看出此时椅子上动态特征点已被完全剔除，并且人周围的静态特征点并没有被后续加上的两种策略暴力剔除，可见本文算法在剔除动态特征点的同时保留了大部分静态特征点，为后续跟踪线程的相机位姿估计提供准确的先验信息。



(a) ORB-SLAM2　(b) Ours-1　(c) Ours-2

图 6　动态特征点剔除效果

Fig. 6　Dynamic feature point rejection effect

3.3　算法性能评估

本文算法在跟踪线程中利用提取到的 ORB 静态特征点对相机位姿进行估计，为了验证本文算法对相机位姿估计的准确性，采用 evo 轨迹评估工具，将本文算法与 ORB-SLAM2 以及其他先进的动态视觉 SLAM 算法进行了轨迹对比评估。选取绝对轨迹误差(absolute trajectory error，ATE)和相对轨迹误差(relative trajectory error，RPE)作为算法性能评价指标。两种误差进一步可细分为误差均方根(RMSE)和标准差(STD)，以便更精确地对本文算法生成的 SLAM 轨迹进行定量分析。ATE 直接衡量了相机位姿估计与真实轨迹之间的偏差，是评估算法精度和全局轨迹一致性的直观指标。而 RPE 侧重于分析在相同时间点上估计位姿与真实位姿之间的相对差异，分析了位姿估计在连续时间内的累积误差特性，涵盖平移和旋转两个维度。为了全面评估 ATE 与 RPE，本文采用了均方根误差(RMSE)和标准差(STD)这两个统计量作为量化指标，以便更精确地验证算法的性能表现。

1)误差结果对比

为了体现出本文算法的性能提升程度，分别使用数据集中 5 组序列将 ORB-SLAM2 算法与本文算法进行对比。定义提升效率 I，公式如下：

$$I = \frac{B - A}{B} \times 100\% \tag{19}$$

在对比分析中，A 代表通过本文改进后的算法所获得的数据；而 B 为改进之前的 ORB-SLAM2 算法数据，为了系统地评估这一改进的效果，本文分别展示了 ATE、平移 RPE 和旋转 RPE 三个关键指标在改进前后的对比结果，这些结果汇总如表 1～3 所示。通过这些表格，可以清晰地观察到改进算法在精度方面的提升。

表 1　绝对轨迹误差(ATE)的评估

Table 1　Evaluation of absolute trajectory error (ATE)

序列	ORB-SLAM2		本文		I/%	
	RMSE	STD	RMSE	STD	RMSE	STD
walking_xyz	0.752 1	0.375 9	0.021 6	0.008 4	97.13	97.76
walking_static	0.390 0	0.160 2	0.006 9	0.003 0	98.23	98.13
walking_rpy	0.870 5	0.452 0	0.028 9	0.015 5	96.68	96.57
walking_half	0.486 3	0.229 0	0.023 4	0.011 3	95.19	95.07
sitting_static	0.008 7	0.004 3	0.005 7	0.002 7	34.48	37.21

表 2　平移相对轨迹误差(RPE)的评估

Table 2　Evaluation of translational relative trajectory error (RPE)

序列	ORB-SLAM2		本文		I/%	
	RMSE	STD	RMSE	STD	RMSE	STD
walking_xyz	0.412 4	0.268 4	0.011 0	0.006 0	97.33	97.76
walking_static	0.216 2	0.196 2	0.005 7	0.002 8	97.36	98.57
walking_rpy	0.424 9	0.316 6	0.021 1	0.012 6	95.03	96.02
walking_half	0.355 0	0.281 0	0.012 4	0.006 7	96.51	96.52
sitting_static	0.009 5	0.004 6	0.004 7	0.002 2	50.53	51.06

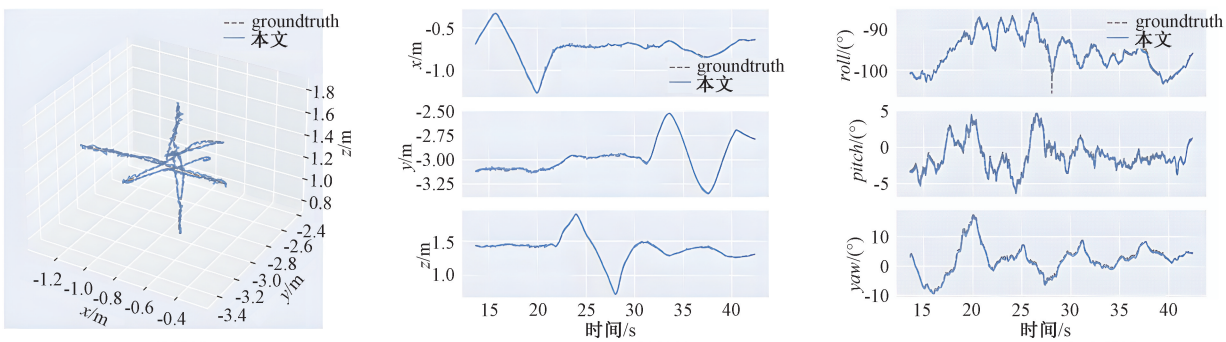
表 3　旋转相对轨迹误差(RPE)的评估

Table 3　Evaluation of rotational relative trajectory error (RPE)

序列	ORB-SLAM2		本文		I/%	
	RMSE	STD	RMSE	STD	RMSE	STD
walking_xyz	7.743 2	4.989 5	0.384 9	0.271 3	95.03	94.56
walking_static	3.895 8	3.509 5	0.168 4	0.084 8	95.68	97.58
walking_rpy	8.080 2	5.949 9	0.654 3	0.397 4	91.90	93.32
walking_half	7.374 4	5.755 8	0.392 4	0.204 4	94.68	96.45
sitting_static	0.288 1	0.124 4	0.155 1	0.082 2	46.16	33.92

分析上述表 1～3 可以看出在绝对轨迹误差(ATE)的评估下,针对 4 组高动态序列,经过改进的算法在均方根误差(RMSE)、以及标准差(STD)方面相较于原 ORB-SLAM2 算法有着显著的优化。具体来说,误差的降低幅度均达到了 95%以上,这一显著的改进表明了本文算法在应对高动态场景时,对于提升视觉 SLAM 系统的定位精度和鲁棒性具有显著效果。特别地,在 walking_static 序列中,本文算法在 RMSE 和 STD 这两个指标上分别实现了高达 98.23%和 98.13%的提升,这进一步证实了本文算法在高动态环境中的出色性能。

然而,当使用低动态序列 sitting_static 进行测试时,虽然本文算法也在各项指标上有所提升,但这种提升相较于高动态序列而言并不那么显著。这主要是因为该序列中的动态物体较少,且原 ORB-SLAM2 算法在低动态环境下已经具备相当好的性能。从对表 2 和表 3 的细致对比中,可见无论所示 RPE 的平移部分还是旋转部分,在高动态序列中,本文算法表现出了明显的性能提升;而在低动态序列中,虽然有所提升,但提升幅度并不显著。这一结果进一步验证了本文算法在处理高动态场景时的优越性和在低动态场景下的稳定性。本文算法定位轨迹如图 7 所示,轨迹误差图如图 8 所示。



(a) 三维轨迹
(a) 3D trajectory

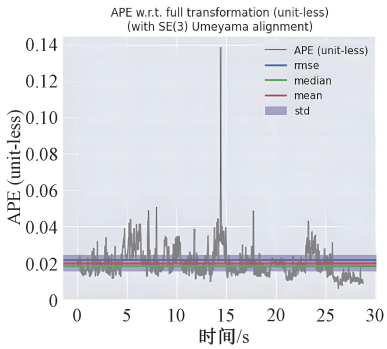
(b) x、y、z三轴轨迹
(b) x、y、z axis trajectories

(c) r、p、y三角度轨迹
(c) r、p、y angle trajectories

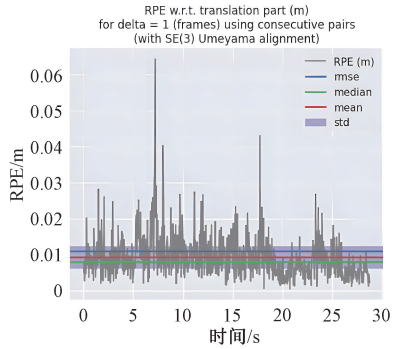
图 7　本文算法定位轨迹

Fig. 7　The algorithm in this paper localizes the trajectory

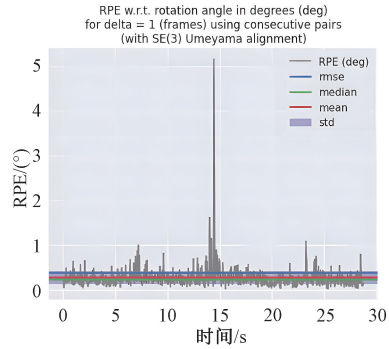
图 7(a)～(c)展示了真实轨迹和本文算法估计的轨迹在 walking_xyz 序列上的轨迹对比,图中所展示的虚线是真实轨迹,蓝色实线是本算法估计出的轨迹。具体地,图 7(b)为相机运动轨迹分别在 x、y、z 三轴上的轨迹对比;图 7(c)为相机运动轨迹分别在滚转角 roll、俯仰角 pitch、偏航角 yaw 三个角度上的轨迹对比。可以看出无论



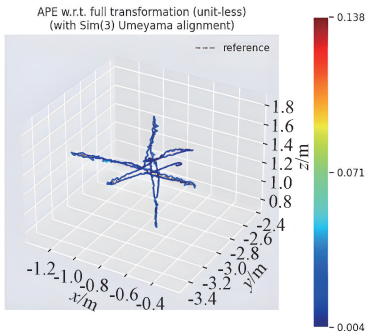
(a) ATE误差数值图
(a) Plot of ATE error values



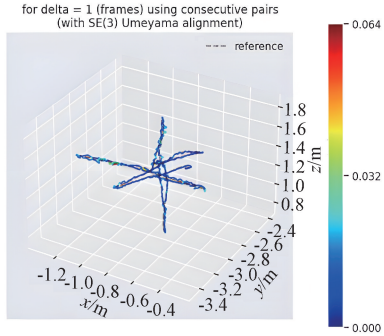
(b) RPE平移部分误差数值图
(b) Plot of RPE translation section error values



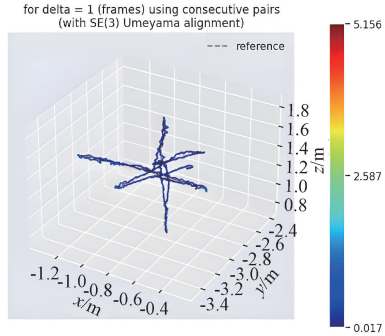
(c) RPE旋转部分误差数值图
(c) Plot of RPE rotating part error values



(d) ATE误差轨迹图
(d) ATE error trajectory diagram



(e) RPE平移部分误差轨迹图
(e) RPE translation part of the error trajectory map



(f) RPE旋转部分误差轨迹图
(f) Error trajectory of the rotating part of the RPE

图 8　轨迹误差图

Fig. 8　Trajectory error diagram

是在三轴上还是在三角度上真实轨迹与本文算法估计出的轨迹几乎重合，证明了本文算法的准确性和鲁棒性。图 8(a)～(c)展示了在 walking_xyz 序列上 ATE 和 RPE 各项指标的数值图。图 8(d)～(f)展示了在 walking_xyz 序列上 ATE 和 RPE 轨迹误差的可视化图。线条的颜色从蓝色到红色表示轨迹误差逐渐升高，图中可见大部分线段颜色为蓝色，验证了本文算法估计的轨迹的误差在数据集序列全时间段都保持在低数值和稳定的状态，进一步说明了系统的稳定性与鲁棒性。因此经过实验验证，本文算法在动态视觉 SLAM 领域有着较高的计算精度以及较强的鲁棒性。

2)与同类型其他算法对比

为了进一步验证本文算法的性能，将其与同类型的其他动态视觉 SLAM 算法进行对比，选取 DS-SLAM、Dyna-SLAM、RDMO-SLAM、RDS-SLAM 四种算法与本文算法进行定量分析，本文算法与同类型算法的 ATE、平移 RPE 及旋转 RPE 对比结果如表 4～6 所示。表中标有下划线的数据为次优效果数据，加粗的数据为最优效果数据。从表 4 看出，在 walking_xyz 和 walking_static 数据集序列上，Dyna-SLAM 与本文算法在 ATE 的 RMSE 指标上分别相差 0.005 2、0.000 1，改进的本文算法误差略高于 Dyna-SLAM；对于平移 RPE，表 5 的数据展示了相较于同类型其他算法，本文算法无论在 RMSE 还是 STD 指标上都是最低值，即最低的误差和最高的定位精度。表 6 展示了在 5 组序列中各算法旋转 RPE 的对比情况，可以看出本文算法的定位误差远小于 RDMO-SLAM、RDS-SLAM 的误差。综上所述，本文算法在各个数据集序列以及评价指标上都表现出最优结果，充分证明了其具有较高的定位精度。

3.4　地图构建

八叉树地图(Octomap)和语义对象度量地图(semantic object metric Map)是移动机器人领域中两种重要的地图表示方法，它们分别提供了对机器人周围环境的不同层次的理解。

1)语义对象度量地图构建

图 9 是 TUM RGB-D 数据集的 walking_rpy 序列中的语义对象度量地图。图中语义地图给出了相机在场景中某一时刻的位姿、采集到的语义对象、语义 ID 以及空间坐标位置(红色方块表示显示器，紫色方块代表椅子)，其坐标是相对于第一帧图像坐标系转换而来的。

2)八叉树地图构建

八叉树地图是一种用于表示三维空间信息的概率地图。它将空间划分为一系列八叉树结构的体素(voxel)，每个体素表示一定大小的三维空间。每个体素存储了一个

表 4　绝对轨迹误差(ATE)的评估(同类型算法)

Table 4　Evaluation of absolute trajectory error (ATE) (homogeneous algorithm)

序列	DS-SLAM		Dyna-SLAM		RDMO-SLAM		RDS-SLAM		本文	
	RMSE	STD	RMSE	STD	RMSE	STD	RMSE	STD	RMSE	STD
walking_xyz	0.024 7	0.016 1	0.016 4	0.008 6	0.022 6	0.013 7	0.028 1	0.016 7	0.021 6	**0.008 4**
walking_static	0.008 1	0.003 3	0.006 8	0.003 2	0.012 6	0.007 1	0.041 9	0.034 8	0.006 9	**0.003 0**
walking_rpy	0.444 2	0.235 0	0.035 4	0.019 0	0.128 3	0.104 7	0.111 4	0.092 0	**0.028 9**	**0.015 5**
walking_half	0.030 3	0.015 9	0.029 6	0.015 7	0.030 4	0.014 1	0.028 2	0.015 5	**0.023 4**	**0.011 3**
sitting_static	0.006 5	0.003 3	0.010 8	0.005 6	0.006 6	0.003 3	0.010 7	0.005 0	**0.005 7**	**0.002 7**

表 5　平移相对轨迹误差(RPE)的评估(同类型算法)

Table 5　Evaluation of translational relative trajectory error (RPE) (homogeneous algorithm)

序列	DS-SLAM		Dyna-SLAM		RDMO-SLAM		RDS-SLAM		本文	
	RMSE	STD	RMSE	STD	RMSE	STD	RMSE	STD	RMSE	STD
walking_xyz	0.033 3	0.022 9	0.021 7	0.011 9	0.029 9	0.018 8	0.028 1	0.016 7	**0.011 0**	**0.006 0**
walking_static	0.010 2	0.003 8	0.008 9	0.004 4	0.016 0	0.009 0	0.041 9	0.034 8	**0.005 7**	**0.002 8**
walking_rpy	0.150 3	0.116 8	0.044 8	0.026 2	0.139 6	0.117 6	0.111 4	0.092 0	**0.021 1**	**0.012 6**
walking_half	0.029 7	0.015 2	0.028 4	0.014 9	0.029 4	0.013 0	0.028 2	0.015 5	**0.012 4**	**0.006 7**
sitting_static	0.007 8	0.003 8	0.012 6	0.006 7	0.009 0	0.004 0	0.010 7	0.005 0	**0.004 7**	**0.002 2**

表 6　旋转相对轨迹误差(RPE)的评估(同类型算法)

Table 6　Evaluation of rotational relative trajectory error (RPE) (homogeneous algorithm)

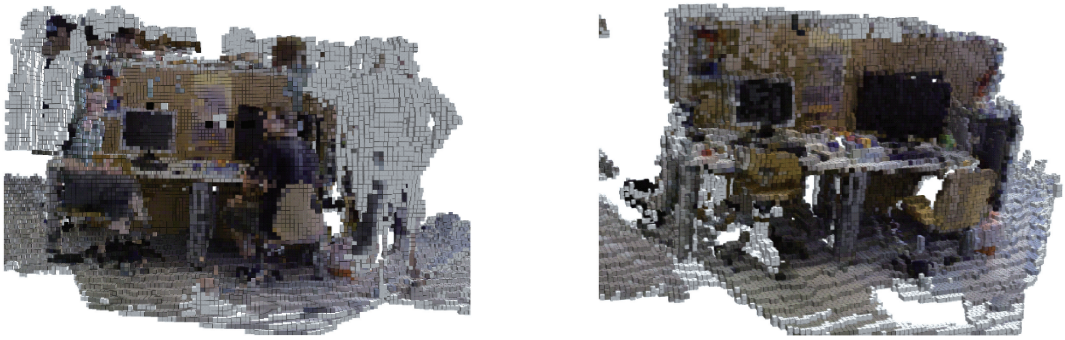
序列	DS-SLAM		Dyna-SLAM		RDMO-SLAM		RDS-SLAM		本文	
	RMSE	STD	RMSE	STD	RMSE	STD	RMSE	STD	RMSE	STD
walking_xyz	0.826 6	0.282 6	0.628 4	0.384 8	0.799 0	0.550 2	0.723 6	0.443 5	**0.384 9**	**0.271 3**
walking_static	0.269 0	0.121 5	0.261 2	0.125 9	0.338 5	0.161 2	1.168 6	0.991 7	**0.168 4**	**0.084 8**
walking_rpy	3.004 2	2.306 5	0.989 4	0.570 1	2.547 2	2.060 7	9.319 2	8.572 0	**0.654 3**	**0.397 4**
walking_half	0.814 2	0.410 1	0.784 2	0.401 2	0.791 5	0.378 2	0.821 6	0.434 7	**0.392 4**	**0.204 4**
sitting_static	0.273 5	0.121 5	0.341 6	0.164 2	0.291 0	0.133 0	0.309 1	0.132 5	**0.155 1**	**0.082 2**

图 9　walking_rpy 序列语义对象度量地图

Fig. 9　Walking_rpy sequence semantic object metrics map

概率值,表示该空间被占据(occupied)、空闲(free)或未知(unknown)的可能性。这种表示方法允许机器人以较高的精度和效率构建和更新三维地图,对于导航和避障任务非常有用。

图 10 为在 walking_rpy 序列下对场景进行八叉树地图构建的结果。图 10(a)为系统在不剔除动态特征点的情况下构建的地图,可以看出移动的人这样的动态因素严重影响了建模的效果。而图 10(b)则展示了剔除动态特征点后构建的地图效果,可见排除了动态干扰后可以更为准确地对场景进行重建。



(a) 有动态特征
(a) With dynamic features

(b) 消除动态特征
(b) Elimination of dynamic features

图 10　walking_rpy 序列八叉树地图构建

Fig. 10　Walking_rpy sequential octree map construction

4　结　　论

提出的基于全景分割和多视图几何方法的动态视觉 SLAM 算法能够准确地识别出场景中任何物体，对于动态物体(包括先验、平行、潜在的动态物)可以实现全面的筛查与剔除，排除动态物体对系统的干扰，并尽可能地保留更多静态特征，提高了视觉 SLAM 的准确性与鲁棒性。本研究实验结果表明，与 ORB-SLAM2 算法相比，本文算法在高动态序列下能够大幅度地提升视觉 SLAM 定位的精确性与鲁棒性，与同类型的先进算法比较，本文算法在定位精度上有小幅提升。消除动态物干扰后的建图效果显著，高度准确地还原了任务场景。因此本文算法在动态视觉 SLAM 领域具有较好的表现，有望未来在无人机领域扮演更加重要的角色。但是，仍然存在一些需要改进的地方，例如，需要进一步提高算法运行速度，优化算法的实时性，并需要构建质量更高的八叉树地图，以便用于无人机视觉导航等领域。

参考文献

[1] 田野，陈宏巍，王法胜，等. 室内移动机器人的 SLAM 算法综述[J]. 计算机科学，2021，48(9)：223-234.
TIAN Y，CHEN H W，WANG F SH，et al. A review of SLAM algorithms for indoor mobile robots[J]. Computer Science，2021，48(9)：223-234.

[2] XIAO L H，WANG J G，QIU X S，et al. Dynamic-SLAM：Semantic monocular visual localization and mapping based on deep learning in dynamic environment[J]. Robotics and Autonomous Systems，2019，117：1-16.

[3] KOSTAVELIS I，GARATERATOS A. Semantic mapping for mobile robotics tasks：A survey[J]. Robotics and Autonomous Systems，2015，66：86-103.

[4] YU CH，LIU Z，LIU X，et al. DS-SLAM：A semantic visual slam towards dynamic environments[C]. 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems(IROS). IEEE，2018：1168-1174.

[5] BESCÓS B，FÁCIL J M，CIVERA J，et al. DynaSLAM：Tracking，mapping and inpainting in dynamic scenes[J]. IEEE Robotics and Automation Letters，2018，3(4)：1-1.

[6] CUI L Y，MA CH W. SOF-SLAM：A semantic visual SLAM for dynamic environments[J]. IEEE Access，2019，7：166528-166539.

[7] 马萍. 复杂环境中联合 CNN 与 IMU 的单目视觉 SLAM 方法研究[D]. 北京：中国科学院大学(中国科学院长春光学精密机械与物理研究所)，2020.
MA P. Research on monocular vision SLAM method with joint CNN and IMU in complex environment[D]. Beijing：University of Chinese Academy of Sciences (Changchun Institute of Optical Precision Machinery and Physics，Chinese Academy of Sciences)，2020.

[8] SONG S，LIM H，LEE A J，et al. DynaVINS：A visual-inertial SLAM for dynamic environments[J]. IEEE Robotics and Automation Letters，2022，7(4)：11523-11530.

[9] LIU Y B，MIURA J. RDS-SLAM：Real-time dynamic SLAM using semantic segmentation methods[J]. IEEE Access，2021，9：23772-23785.

[10] QIAN J，CHATRATH V，YANG J，et al. POCD：Probabilistic object-level change detection and volumetric mapping in semi-static scenes[J]. Robotics：Science and Systems，2022，5：11573-11588.

[11] LIU Y B，MIURA J. RDMO-SLAM：Real-time visual SLAM for dynamic environments using semantic label prediction with optical flow[J]. IEEE Access，2021，9：106981-106997.

[12] 马哲伟，周福强，王少红. 昏暗环境下自适应 ORB-

SLAM2 算法研究[J]. 电子测量技术,2024,47(6):94-99.
MA ZH W,ZHOU F Q,WANG SH H. Research on Adaptive ORB-SLAM2 Algorithm in Dim Environment [J]. Electronic Measurement Technology, 2024,47(6):94-99.

[13] 谢波,张国良,李歆,等.一种单目 VIO 定位精度与跟踪稳定性优化方法[J].国外电子测量技术,2023,42(4):23-30.
XIE B, ZHANG G L, LI X, et al. Optimization method for positioning accuracy and tracking stability of monocular VIO [J]. Foreign Electronic Technology,2023,42(4): 23-30.

[14] 张耀,吴一全,陈慧娴.基于深度学习的视觉同时定位与建图研究进展[J].仪器仪表学报,2023,44(7):214-241.
ZHANG Y, WU Y Q, CHEN H X. Research progress of visual simultaneous localization and mapping based on deep learning[J]. Chinese Journal of Scientific Instrument, 2023,44(7): 214-241.

[15] PAGAD S, ARARWAL D, NARAYANAN S, et al. Robust method for removing dynamic objects from point clouds[C]. 2020 IEEE International Conference on Robotics and Automation (ICRA) IEEE, 2020: 10765-10771.

[16] ZELLER N, QUINT F, STILLA U. From the calibration of a light-field camera to direct plenoptic odometry[J]. IEEE Journal of Selected Topics in Signal Processing,2017,11(7): 1004-1019.

[17] RUS R B, COUSINS S. 3D is here: Point cloud library(pcl)[C]. 2011 IEEE International Conference on Robotics and Automation, 2011: 1-4.

[18] HORNUNG A, WURM K M, BENNEWITZ M, et al. OctoMap: An efficient probabilistic 3D mapping framework based on octrees[J]. Autonomous Robots, 2013, 34: 189-206.

[19] LI G, YU L, FEI SH. A binocular MSCKF-based visual inertial odometry system using LK optical flow[J]. Journal of Intelligent & Robotic Systems, 2020, 100(3): 1179-1194.

[20] STURM J, ENGELHARD N, ENDRES F, et al. A benchmark for the evaluation of RGB-D SLAM systems[C]. 2012 IEEE/RSJ International Conference on Intelligent Robots and Systems, IEEE, 2012: 573-580.

作者简介

王爽,硕士研究生,主要研究方向为机器人视觉 SLAM。
E-mail:w2380694886@163.com

刘云平(通信作者),教授,博士生导师,博士,主要研究方向为机器人技术与智能装备。
E-mail:002105@nuist.edu.cn

张柄棋,硕士研究生,主要研究方向为机器人路径规划。
E-mail:zhangbq1822@163.com

陆旭春,硕士研究生,主要研究方向为机器人目标跟踪。
E-mail:1073979365@qq.com

徐梁,教授,博士生导师,博士,主要研究方向为无人机集群。
E-mail:3930997927@qq.com